The Plant Cell, Vol. 5, 1401–1410, October 1993

Gene Expression during Plant Embryogenesis and Germination: An Overview

Terry L. Thomas

Department of Biology, Texas A&M University, College Station, Texas 77843

INTRODUCTION

Seed development represents a unique transition state in the life cycle of higher plants, providing the physical link between parental and progeny sporophytic generations. During embryogenesis, the root and shoot apical meristems are specified, thus establishing the basic architecture of the seedling, and differentiation of vegetative tissue and organ systems occurs. Maturation events prepare the seed for germination and subsequent development of the mature plant. During maturation, the developing seed increases dramatically in volume and mass due to significant cell expansion and the concomitant accumulation and storage of protein and lipid to be used as nitrogen and carbon sources during germination. Early during the maturation phase, abscisic acid (ABA) levels reach a maximum, suppressing precocious germination and modulating gene expression. Desiccation ensures seed dormancy even in the absence of ABA and serves as the boundary between seed maturation and germination. Following imbibition, germination and seedling growth ensue, driven metabolically by the hydrolysis of protein and lipid stored during maturation.

This review will provide an eclectic overview of gene expression in angiosperm embryogenesis and germination. I will focus primarily on dicot plants and will address only a small number of the more than 3×10^4 different genes expressed in embryos and seedlings. First, gene expression in early embryogenesis will be reviewed, focusing on selected results obtained with carrot somatic embryos. Control of gene expression during seed maturation will then be discussed, with an emphasis on *cis*- and *trans*-acting factors involved in regulating maturation phase genes. A brief description of gene expression during germination and the relationship of embryonic and vegetative gene expression programs will follow. The review will conclude with a prospectus for the study of global gene expression patterns in early embryos.

GENE EXPRESSION IN EARLY EMBRYOS

Although most of the major discernible morphogenetic events in plants occur after germination, the overall architectural pattern of the mature plant is established during embryogenesis (see West and Harada, 1993, this issue). Following fertilization, an asymmetric division of the zygote of many dicots yields a basal cell that will produce the suspensor and an apical cell—the first embryonic cell. Ensuing cleavages of the apical cell yield a radially symmetric globular embryo with a differentiated protoderm, or dermatogen. Further cell divisions and subsequent morphogenesis break radial symmetry, yielding the bilaterally symmetric, heart stage embryo with root and shoot apices, incipient cotyledons, and provascular tissue (for recent reviews, see Meinke, 1991; Van Engelen and De Vries, 1992; De Jong et al., 1993a; West and Harada, 1993, this issue). So far, few genes have been identified that are expressed specifically during early embryogenesis, primarily because of technical difficulties associated with the mass ratios of the embryo and the surrounding maternal tissue and the lack of molecular and cellular markers to direct screening efforts. The most significant progress in understanding the molecular and cellular events of early embryogenesis has resulted from experiments on somatic embryos of a few plant species (De Jong et al., 1993a; Liu et al., 1993; Zimmerman, 1993, this issue). In this regard, analysis of embryogenic carrot cell cultures has proven most productive (Van Engelen and De Vries, 1992; De Jong et al., 1993a).

Several genes have been identified from carrot suspension cultures that encode glycosylated extracellular proteins (EPs) secreted into the medium. For example, EP1 is closely related to *Brassica* and Arabidopsis *S* glycoproteins and to the putative *S*-like receptor protein kinases from maize and Arabidopsis (see Nasrallah and Nasrallah, 1993, this issue). EP1 mRNA accumulates in the inner and outer integument of the epidermis of developing seeds and in nonembryogenic cells in carrot cultures (Van Engelen et al., 1993). Although EP1 is not expressed early in somatic embryos, it is interesting because of the potential relationship of EP1 congeners to extracellular serine/threonine protein kinases, which might be involved in embryo-related signal transduction. EP2 encodes an extracellular lipid transfer protein and is one of the few characterized genes expressed in preglobular somatic and zygotic embryos (Sterk et al., 1991). EP2 mRNA accumulates specifically in cells of the protoderm, where it is presumed to be involved in cutinization of the embryo surface.

EP3 is a 32-kD extracellular acidic endochitinase that can rescue a temperature-sensitive mutation (*ts11*) that arrests

embryogenesis at the preglobular stage (De Jong et al., 1992). In *ts11* somatic embryos, protoderm development and expression of EP2 in the protoderm are disrupted at the restrictive temperature. The addition of EP3 to the medium restores temperature-sensitive protoderm development, as determined by the recovery of EP2 mRNA expression in the protoderm (De Jong et al., 1993b). Surprisingly, EP3 rescue of *ts11* embryos at the restrictive temperature is mimicked by the addition of a *Rhizobium* nodulation (Nod) factor, an *N*-acetylglucosamine–containing lipooligosaccharide; these results suggest that the endochitinase may generate plant analogs of Nod factors (De Jong et al., 1993b). Another enzyme secreted into the medium, a cationic peroxidase, can restore globular embryo development in carrot cultures in which somatic embryogenesis is impaired by tunicamycin (Cordewener et al., 1991). These genes, like others reviewed elsewhere, will serve as important tools to further probe developmental events in the early embryo. For example, the use of EP2 gene expression as a determinant of protodermal differentiation has already proven useful in the study of the *ts11* mutant (De Jong et al., 1993b), and similar studies with Arabidopsis embryo-defective and pattern mutants are anticipated (S. De Vries, personal communication).

CONTROL OF GENE EXPRESSION DURING SEED MATURATION

Morphologically and biochemically, the most dramatic events of seed development occur during maturation. These include rapid expansion of the cotyledons and concomitant synthesis and accumulation of proteins, lipids, and, in some cases, carbohydrates. Maturation also involves significant physiological adaptations that occur within the seed to ensure embryo dormancy, including accumulation of ABA. During late embryogenesis, ABA prevents precocious germination prior to desiccation and promotes embryo maturation (reviewed in Quatrano, 1987). The events of maturation require the coordinated expression of a number of genes in response to ABA.

A Simple Model of Seed Protein and *late embryo abundant (lea)* Class Gene Promoters

Previously, we suggested a bipartite organization of seed protein and *lea* class gene regulatory ensembles (Thomas et al., 1991). In this model, proximal promoter regions specify seed expression and more distal regions function to refine and enhance the basic expression patterns conferred by proximal regions. The bipartite model was developed for seed protein and *lea* class genes but may apply to other highly expressed, tissue-specific genes.

Figures 1 and 2 summarize results from the author's laboratory in terms of the bipartite model. Proximal promoter regions

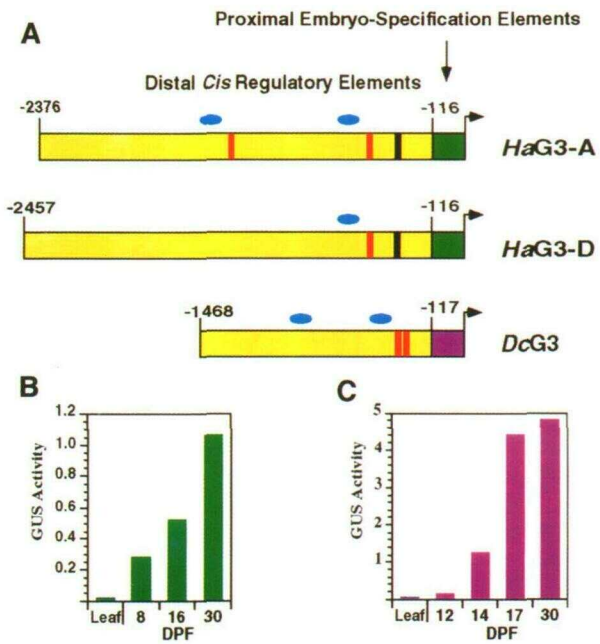

Figure 1. Functional Organization of Seed Protein and *lea* Class Gene Regulatory Ensembles.

(A) Regulatory ensembles of sunflower helianthinin and carrot *Dc*G3 genes. Proximal promoter regions are green and purple for the helianthinin genes and *Dc*G3 gene, respectively, and are indicated by the vertical arrow; distal regulatory regions are in yellow. Black rectangle indicates the E-box motif (CANNTG); ABREs are indicated by red rectangles. Blue ovals are located above A/T-rich motifs that bind ubiquitous sunflower nuclear proteins that are probably high mobility group–like (HMG-like) chromosomal proteins.

(B) Helianthinin promoter proximal region drives GUS reporter gene expression in transgenic tobacco seeds. Fluorometric GUS analysis of developing seeds and leaves of transgenic tobacco containing the helianthinin promoter proximal region (−116 to +24) fused to a translationally enhanced GUS reporter gene. GUS activity in 30-days postflowering (DPF) seeds is ~100 times that in leaves.

(C) *Dc*G3 promoter proximal region drives GUS reporter gene expression in transgenic tobacco seeds. Fluorometric GUS analysis of developing seeds and leaves of transgenic tobacco containing the *Dc*G3 promoter proximal region (−117 to +26) fused to the GUS reporter gene. GUS activity in 30-DPF seeds is ~450 times that in leaves. GUS activity units are pmoles 4-methylumbelliferone/μg/min.

(−116 to +24) of two nonallelic sunflower genes, *Ha*G3-A and *Ha*G3-D, both encoding the legumin-like seed protein helianthinin, are identical (Figure 1A). This region can direct seed-specific expression of the β-glucuronidase (GUS) reporter gene (Figure 1B); expression is localized primarily to the embryonic apical region and the cotyledons (Figure 2A). Distal elements in the helianthinin regulatory ensemble, including A/T-rich enhancer elements and ABA-responsive elements (ABREs), expand the tissue boundaries of helianthinin expression and modulate expression both temporally and

quantitatively (Bogue et al., 1990). Additional elements maintain hierarchical control over distal ABREs, so that in the helianthinin promoters these elements are ABA responsive only in developing seeds (Thomas et al., 1991; A. Nunberg, Z. Li, M. Bogue, J. Vivekananda, A. Reddy, and T.L. Thomas, unpublished results).

The promoter of the carrot *lea* class gene *Dc*G3 has a nearly identical organization (Figure 1A). A proximal region (−117 to +26) confers ABA-independent, seed-specific GUS expression in transgenic tobacco (Figure 1C). The expression pattern is similar, but not identical, to that of the helianthinin promoter proximal region (compare Figures 2A and 2B). Distal ABREs modulate *Dc*G3 expression in response to changing ABA concentrations in developing seeds and in desiccated vegetative tissues (Vivekananda et al., 1992).

The bipartite model applies to other dicot seed protein genes as well. A proximal promoter region of the phaseolin gene, including a portion of the transcriptional activator sequence (UAS1), confers seed-specific GUS expression; distal positive and negative elements provide additional spatial and temporal regulation of phaseolin expression (Bustos et al., 1991). Goldberg et al. (1989) suggested that proximal elements up to −77 of the soybean lectin (*Le1*) and −66 of the glycinin (*Gy1*) genes were required for correct spatial and temporal expression of these genes; more distal elements were again required for their quantitative regulation. Furthermore, proximal promoter regions of the soybean β-conglycinin α′ subunit gene (−140 to +13) and of the *Phaseolus vulgaris* (French bean) lectin (*dlec2*) gene (−125 to +1) are sufficient for tissue-specific expression (Voelker et al., 1987; Lessard et al., 1993).

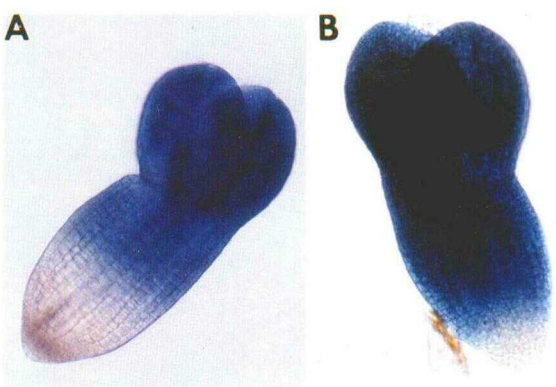

Figure 2. Localization of GUS Expression Driven by Proximal Promoter Regions of Sunflower Helianthinin and Carrot *Dc*G3 Genes in Transgenic Tobacco Seeds.

(A) X-Gluc staining of 18-DPF transgenic tobacco seeds containing the same construction as in Figure 1B.

(B) X-Gluc staining of 18-DPF transgenic tobacco seeds containing the same construction as in Figure 1C.

Seed Protein Gene Regulation

Expression of seed storage proteins is rigorously tissue specific; expression occurs primarily in the embryonic axis and cotyledons and, in some cases, in the endosperm of developing seeds but never in mature vegetative tissues. The expression patterns of seed protein genes are highly regulated spatially and temporally during seed development (Goldberg et al., 1989; Perez-Grau and Goldberg, 1989; Guerche et al., 1990), and resulting storage proteins are processed and targeted to protein bodies (Shotwell and Larkins, 1989). The tissue-specific expression of seed protein genes during seed development has made them especially useful as models of gene expression in embryogenesis, and as a consequence, there have been extensive studies on the structure and expression of these genes from numerous plant species.

Table 1 lists several classes of DNA sequence motifs and, in some cases, the associated *trans*-acting factors that have been implicated in the regulated expression of seed protein genes. Many of these motifs were identified because of their phylogenetic conservation; few have been analyzed in any detail. Furthermore, a number of detailed studies have failed to provide a definitive picture of the roles of specific sequences in seed protein gene expression.

Why has it been so hard to define seed-specific *cis*-regulatory elements? Two major factors have contributed to this conundrum. First, regulatory ensembles of seed protein genes are extensive, often including more than a kilobase of upstream sequence (Figure 1), making it difficult to identify DNA sequences for detailed functional analysis. Second, seed protein gene regulatory ensembles are the result of combinatorial interactions of multiple DNA elements. Therefore, it is not surprising that analysis of individual elements comprising these ensembles has sometimes led to conflicting results on the role of specific elements in a regulatory ensemble (Bäumlein et al., 1991, 1992).

Proximal Seed Specification Elements

The maize regulatory locus *opaque2* (*o2*) plays a critical role in regulating expression of the genes encoding one class of maize seed proteins, the 22-kD zeins (see Lopes and Larkins, 1993, this issue). The O2 protein recognizes the sequence TCCACGTAGA in the promoter of 22-kD zein genes (Schmidt et al., 1992). This motif is present only in promoters of 22-kD zein genes; it is not present in the promoters of other classes of zein genes, although there are closely related sequences. O2, a basic leucine zipper (bZIP) protein, functions in yeast as a *trans*-activator of promoters that include the O2 target site (Schmidt et al., 1992). Recently, a second maize bZIP protein, OHP1, was isolated based on its interaction with O2 (Pysh et al., 1993). OHP1 binds to the O2 target sequence in vitro as a homodimer or as a heterodimer with O2. Because mutations in O2 abolish expression of the 22-kD zeins, the simplest model

Table 1. DNA Sequences Implicated in Regulating Gene Expression in Developing Seeds

Sequence Motif[a]	Gene	Regulatory Function	*trans*-Acting Factor(s)	Reference
ACGT Motifs				
GGACACGTGGC	Wheat Em	ABA regulation	EMBP-1	Guiltinan et al. (1990)
CCGTACGTGGC	Rice *Rab16*	ABA/desiccation	TAF-1	Mundy et al. (1990); Oeda et al. (1991)
TGTTACGTGCC	Carrot *Dc3*	ABA/seed expression	(bZIP)[b]	H. Chung and T.L. Thomas (unpublished results)
TTCCACGTAGA	Maize 22-kD zein	Seed expression	O2	Schmidt et al. (1992)
ACACACGTCAA	French bean phaseolin	Seed expression	O2	M. Bustos (personal communication)
-GCCACGTGGC	Synthetic (PA)[c]	Root expression	TAF-1	Salinas et al. (1992)
--GTACGTGGC	Synthetic (Iwt)[d]	Seed expression	(bZIP)	Salinas et al. (1992)
ATGTACGAAGC	Sunflower helianthinin	ABA regulation	(bZIP)	A. Nunberg, Z. Li, M. Bogue, J. Vivekananda, A. Reddy, and T.L. Thomas (unpublished results)
GGACGCGTGGC	Synthetic (*hex-3*)[e]	Seed/ABA/desiccation	(bZIP)	Lam and Chua (1991)
RY Repeat Motifs				
CATGCATG	*Vicia faba LeB4*	Quantitative seed expression	Unknown	Bäumlein et al. (1992)
	Soybean glycinin (*Gy2*)	Quantitative seed expression	Unknown	Lelievre et al. (1992)
	Other dicot and monocot seed protein genes			Dickinson et al. (1988)
CATGCATGCA (Sph Element)	Maize *C1*	Seed expression/ABA	VP1	Hattori et al. (1992)
	Wheat Em	Unknown	Unknown	Marcotte et al. (1989)
	Maize *Rab17*	Unknown	Unknown	Vilardell et al. (1990)
$(CA)_n$	Many monocot and dicot seed protein genes	(Negative regulatory sequence)	Unknown	Goldberg (1986); Lessard et al. (1993); Vellanoweth and Okita (1993)
E-Box				
CANNTG	French bean phaseolin	(Seed expression)	(bHLH)	Kawagoe and Murai (1992)
A/T Motifs	Most dicot and monocot seed protein genes	Quantitative regulation (Nonspecific enhancer)	HMG or HMG-like	Jordano et al. (1989); Bustos et al. (1989); Pederson et al. (1991); Jofuku et al. (1987); Allen et al. (1989)
WS Motifs				
TGATCT	Sunflower helianthinin	Seed expression	Multiple unknown	A. Nunberg, Z. Li, M. Bogue, J. Vivekananda, A. Reddy, and T.L. Thomas (unpublished results)
AGATGT	Sunflower helianthinin	Seed expression	Multiple unknown	

[a] Abbreviations are as follows: N = A, G, C, or T; R = A or G; Y = C or T; W = A or T; S = G or C.
[b] Parentheses indicate proposed or hypothetical functions or factors.
[c] PA, perfect palindromic sequence containing G-box–related sequence.
[d] Iwt, wild-type motif I containing G-box–related sequence.
[e] *hex-3*, mutant of *hex-1* sequence from wheat histone H3 promoter.

for OHP1 function is as a heterodimeric partner of O2 to activate 22-kD zein genes; however, other models are also under consideration.

The phaseolin UAS1 contains a motif similar to that bound by O2 (Table 1), and, in fact, the phaseolin motif binds O2 synthesized in *Escherichia coli* (M. Bustos, personal communication). However, additional motifs, including an E-box, vicilin box, and putative ABRE, are present in UAS1, and it is likely that more than one sequence motif acts together to mediate seed-specific activation of heterologous promoters by UAS1. Furthermore, the results of Kawagoe and Murai (1992) suggest that an E-box (CANNTG) located at the distal end of UAS1 binds seed nuclear proteins; a functional role for this interaction was hypothesized, but definitive data are not currently available. In animals, this motif frequently binds to basic helix-loop-helix (bHLH) transcription factors and is involved in cell specification (Weintraub et al., 1991). In plants, bHLH proteins are involved in anthocyanin expression in maize seeds and flowers (Ludwig et al., 1989; Goodrich et al., 1992). Therefore, it is possible that a bHLH protein may be involved in the interaction with E-box motifs and may function in the tissue-specific expression of seed protein genes. In this regard, it is noteworthy that the *Dc*G3 proximal seed specification element (−117 to +26) includes three E-box motifs (H. Chung and T.L. Thomas, unpublished results). The proximal promoter regions of soybean and French bean lectin genes have no E-boxes; the latter, however, both contain at least one CATGCCAT motif (Table 1) as does the β-phaseolin gene.

The proximal promoter region (−140 to +13) of the soybean α′ subunit of β-conglycinin is sufficient for seed-specific expression in transgenic tobacco (Lessard et al., 1993). This region contains a binding site for the soybean nuclear factor SEF4 (Lessard et al., 1991) as well as a vicilin box, GCCACCTC (Gatehouse et al., 1986). However, just as with the phaseolin UAS1, additional elements are present in the α′ subunit promoter proximal region that may be involved in seed-specific expression: these include a CATGCAT sequence, an ACGT motif, and an E-box. As a consequence, it is not possible to identify a single motif that confers seed expression of β-conglycinin.

The proximal region of the helianthinin promoter (−116 to +24) contains two copies of the E-box motif. However, mutations in these motifs, including nucleotides required for bHLH binding, do not affect their ability to bind sunflower nuclear proteins, thus indicating that the E-box motifs in the helianthinin promoter proximal region do not function by interacting with bHLH nuclear proteins. Although these results are negative, they are instructive because they illustrate the danger of relying on consensus sequences to identify putative *cis*-regulatory elements. Two sequence motifs, AGATGT and TGATCT, each occur twice in the helianthinin promoter proximal region, AGATGT at −111 and −58 and TGATCT at −83 and −41. These motifs, designated "WS" in Table 1, specifically bind partially purified nuclear proteins from sunflower (Table 1) and appear to be related, having the consensus sequence WGATST, where W = A or T and designates weak (two) hydrogen bonds and S = G or C and designates strong (three) hydrogen bonds.

Disruption of the helianthinin promoter proximal region at −75, including the WS motifs, results in loss of GUS expression or ectopic GUS expression in nonembryonic tissues (A. Nunberg, Z. Li, M. Bogue, J. Vivekananda, A. Reddy, and T.L. Thomas, unpublished results). Point mutations in the WS motifs abolish binding to seed nuclear protein(s); furthermore, these mutations significantly reduce the activity of the promoter in transgenic tobacco (Z. Li and T.L. Thomas, unpublished results). We speculate that the WS motifs act in concert in the function of the proximal promoter region and that an important component of their function is the alteration of the topology of the promoter complex.

Distal Regulatory Elements

Additional elements are required to modulate and expand the dynamic range of seed protein gene expression in time and space (Table 1). The CATGCATG, or RY, motif is widely distributed in dicot and monocot genes and comprises a portion of the 28-bp legumin box (Bäumlein et al., 1986; Dickinson et al., 1988). Deletions of the RY repeat in the *Vicia faba* (broad bean) *LeB4* gene and the soybean glycinin gene result in a decrease in reporter gene expression in transgenic tobacco seeds (Bäumlein et al., 1992; Lelievre et al., 1992). Interactions between the CATGCATG element and more distal sequences are required for high level expression. Furthermore, in the case of *LeB4*, more proximal sequences located 3′ to −68 were implicated in tissue-specific expression. Other distal activation sequences have been described, including UAS2 and UAS3 in the phaseolin gene (Bustos et al., 1991) and the "a" fragment (−257 to −206) of the α′ subunit of the soybean β-conglycinin gene (Lessard et al., 1993). It is instructive to note that closer analysis of a previously reported seed-specific enhancer, which includes the AVCCCA motif (where V = A, C, or G) that binds to another soybean nuclear factor, SEF3 (Chen et al., 1988; Allen et al., 1989), failed to confirm seed-specific enhancement by this motif (Lessard et al., 1993). This demonstrates the need to verify by more than one approach the function of putative *cis*-regulatory elements.

ABA is involved in regulating expression of specific genes in embryos of all seed plants studied so far, including cotton, rape, soybean, sunflower, wheat, rice, barley, and maize (Skriver and Mundy, 1990; Thomas et al., 1991). ABA is required for the accumulation of the 12S seed protein cruciferin in *Brassica* embryos (Finkelstein et al., 1985), for the continued synthesis of the β subunit of β-conglycinin in soybean cotyledons (Bray and Beachy, 1985), and the modulation of sunflower helianthinin gene expression (Thomas et al., 1991; A. Nunberg, Z. Li, M. Bogue, J. Vivekananda, A. Reddy, and T.L. Thomas, unpublished results). In maize, ABA is required for the initiation of synthesis and accumulation of storage globulins (Rivin and Grudt, 1991). Little is known about the sequences involved

in ABA-responsive expression of seed protein genes, the only examples being those of the sunflower helianthinin genes (Table 1). It is interesting that helianthinin ABREs are closely linked to A/T-rich putative enhancer elements (Figure 1); this may account for the exceptionally high levels of ABA induction of the sunflower ABREs compared to ABREs in *lea* class genes (Marcotte et al., 1989; Thomas et al., 1991).

A/T-rich domains, which bind to ubiquitous, prevalent nuclear proteins, are present in the regulatory ensembles of most, if not all, seed protein genes (Table 1). For example, phaseolin UAS3 contains an A/T-rich domain (Bustos et al., 1991) that acts as an enhancer in transgenic tobacco (Bustos et al., 1989). Similarly, the A/T-rich domains of sunflower helianthinin genes drive GUS expression from a truncated cauliflower mosaic virus 35S promoter in seeds of transgenic tobacco (Jordano et al., 1989). In some cases, the nuclear proteins that bind to A/T-rich domains have been shown to be high mobility group chromosomal proteins (Pederson et al., 1991). It is possible that A/T-rich domains act as nonspecific enhancers by functioning as scaffold attachment regions creating transcriptionally active domains that include seed protein and other highly expressed genes.

In addition to these positive regulatory elements, several groups have identified distal negative regulatory elements. The motif AGAAMA, where M = A or C, occurs frequently in negative regulatory sequences such as NRS1 (−391 to −295) and NRS2 (−518 to −418) in the phaseolin gene (Bustos et al., 1991) and may bind the seed nuclear factor AG-1 (Kawagoe and Murai, 1992). The "b" fragment (−206 to −140) of the soybean β-conglycinin gene contains a potent negative element; the motif $(CA)_n$ may play a role in its function (Lessard et al., 1993). This motif is *cis*-dominant to more proximal seed specification elements, but the more distal "a" fragment activation sequence (see above) completely overrides the negative effect of the "b" fragment.

Regulation of *lea* Genes

lea genes are ubiquitous in monocots and dicots (Dure et al., 1989). They are characterized by their ABA-modulated expression in late maturation phase embryos and by their ABA-induced expression in response to various environmental stresses, including desiccation (Skriver and Mundy, 1990). Lea proteins may function to protect cellular structures in mature embryos during seed desiccation and in water-deficient vegetative tissues; however, a direct relationship between these proteins and desiccation tolerance has yet to be demonstrated.

Significant progress has been made in the elucidation of a major class of *cis*-acting DNA sequences involved in the regulated expression of *lea* genes (Table 1). Marcotte et al. (1989) identified an ABRE in the 5′ sequence of the wheat Em gene that is required for ABA-responsive GUS reporter gene expression in rice protoplasts and in transgenic tobacco. The ABRE includes the motif CACGTGGC, which binds to the bZIP transcription factor EMBP-1 (Guiltinan et al., 1990). Other ABREs from *lea* class genes of dicots and monocots are congeners of the wheat ABRE (Table 1); the core of the ABRE consensus sequence is the ACGT motif that in plants has frequently been associated with binding to bZIP transcription factors (Schindler et al., 1992). However, it is apparent that motifs divergent from the CACGTGGC motif, and even from the ACGT core, can confer ABA-responsive gene expression under some circumstances (Table 1).

The *Sph* element, so designated because it contains an *SphI* restriction site, is present in several cereal genes, most notably the maize *C1* gene (Hattori et al., 1992); it represents a more extensive RY repeat (Table 1). The *Sph* element appears to be involved in ABA modulation of *C1* gene expression during seed maturation, and the *Viviparous-1* (*Vp1*) gene product is apparently involved. However, although VP1 behaves as a *trans*-acting factor, VP1 has not so far been shown to bind to the *Sph* element or to any other DNA sequence. VP1 may act indirectly on the *Sph* element via other transcription factors to confer ABA-responsive *C1* expression during seed maturation.

GERMINATION GENE EXPRESSION PROGRAMS

Germination and postgerminative development constitute a critical node in the sporophytic life cycle. During this period, the basic architecture of the mature plant is established, following the pattern formed during embryogenesis, and dramatic physiological changes occur. Lipid and protein reserves are mobilized in the germinating seed to provide carbon and nitrogen for the seedling prior to the initiation of photosynthesis. Distinct mRNA populations are expressed in seedlings as compared to embryos and mature plants, but a number of genes expressed at high levels in seedlings are also represented in developing seeds (Harada et al., 1988; Hughes and Galau, 1988). Four genes, those encoding isocitrate lyase (IL) and malate synthase (MS) as well as two genes whose functions are unknown, are exemplary of germination and postgerminative gene expression programs.

IL and MS are important glyoxylate cycle enzymes involved in mobilizing stored lipids in plant seedlings. The expression of genes encoding these enzymes has been studied in a number of plants, including sunflower, cucumber, and oilseed rape (Allen et al., 1988; Comai et al., 1989; Graham et al., 1990). IL and MS are expressed coordinately at high levels in oilseed rape seedlings; there is significant but lower expression in developing oilseed rape embryos. This pattern is seen with a cucumber MS gene (Graham et al., 1990) and has been inferred for the sunflower IL gene (Allen et al., 1988). In oilseed rape seedlings, IL and MS mRNAs accumulate primarily in the axis and cotyledons; in late postgerminative seedlings, IL transcripts accumulate to high levels throughout cotyledon parenchyma cells (Comai et al., 1989; Dietrich et al., 1989). Thus, not surprisingly, the accumulation patterns of IL and MS

transcripts parallel the lipid mobilization patterns within the cotyledon.

COT44, which probably encodes a cysteine protease, is expressed in the vascular tissue and epidermis of cotyledons but not in storage parenchyma cells of oilseed rape seedlings (Dietrich et al., 1989). This suggests that if COT44 does encode a protease, it probably does not function to mobilize seed storage proteins. On the contrary, epidermal expression of COT44 suggests a potential role in pathogen defense. No information is available on potential *cis*-regulatory elements that may direct expression of COT44 or the IL and MS genes.

The expression pattern of the postgermination abundant gene *AX92* is similar, but not identical, to those of the genes discussed above. *AX92* mRNA is expressed in the ground meristem and mature cortex of the root apex of oilseed rape seedlings and in cortex cells at the root apex of the embryo (Dietrich et al., 1989). Studies on the expression of chimeric GUS reporter genes containing 5′ and 3′ DNA elements of *AX92* showed that 3′ *AX92* elements are required for appropriate quantitative and, more importantly, qualitative control of reporter gene expression (Dietrich et al., 1992). Deletions to −180 had a quantitative effect, but little qualitative effect, on axis expression. However, an element responsible for activating expression in seedling hypocotyl is located between −2400 and −960. Sequences located 3′ to the *AX92* coding region are required for embryonic and seedling axis expression. So far, little is known about the specific 3′ sequences involved or whether regulation is transcriptional or post-transcriptional.

PROSPECTUS

We are on our way to understanding the regulation of gene expression in the maturation phase of seed development, and we will soon understand the specific interactions between *cis*- and *trans*-acting factors required for spatial, temporal, and hormonal regulation of seed protein and *lea* class genes. In some cases, genes encoding the *trans*-acting factors are in hand; others will be available in the near future. With tools such as seed protein and *lea* class *cis*-regulatory elements, their cognate *trans*-acting factors, and the genes encoding important *trans*-acting factors, we can "bootstrap" our way to a second level of the embryo gene regulatory hierarchy. This will provide one approach to understanding regulatory events of early embryogenesis. Even so, this will only be complementary to more direct efforts aimed at understanding the cellular and molecular events of early embryogenesis, a period that is arguably the most important of plant development. In the following, a brief prospectus for the molecular and cellular analysis of early plant embryogenesis is presented.

Kamalay and Goldberg (1980) determined that $\sim 2 \times 10^4$ diverse sequences are expressed in tobacco embryos; this represents an upper bound on the total number of sequences involved in embryogenesis. These studies, however, were insensitive to very early stages of embryogenesis. New technologies, including polymerase chain reaction–based subtractive hybridization (Wang and Brown, 1991), differential display of mRNA (Liang and Pardee, 1992), transposon tagging (Chuck et al., 1993; Fedoroff and Smith, 1993), and amplified antisense RNA (Van Gelder et al., 1990), now allow us to address the complexity of sequences expressed in early embryos, where mRNA mass is limiting, and to isolate genes expressed in the early embryo (Meinke, 1991). Microsurgical (Schiavone and Racusen, 1991) and in situ techniques, in some cases coupled with the previously mentioned techniques, will provide critical new information on early embryo gene expression in time and space.

Assessing the role of newly identified genes expressed in early plant embryos is challenging. So far, stable knockouts of specific genes by homologous recombination, insertional mutagenesis (Meinke, 1991), or site-specific cleavage (Mergny et al., 1992) are problematic. However, antisense (McGurl et al., 1992) and dominant negative (Amaya et al., 1991) knockouts offer attractive alternatives. Overexpression and ectopic expression of wild-type and mutant genes represent another approach. Genes can be overexpressed in appropriate time and space when driven by enhanced, homologous promoters, ectopically expressed in tissues or developmental stages in which the gene would not normally be expressed using the numerous heterologous tissue-specific promoters now available, transiently induced with small molecules such as tetracycline (Gatz et al., 1991), or repressed reversibly with the *lac*/IPTG system (Wilde et al., 1992). Cell and tissue ablation with dominant "killer" genes, such as diphtheria toxin or ribonuclease, driven by cell- and tissue-specific embryo promoters (e.g., Mariani et al., 1990, and references therein) will provide further insight into early embryo gene function. Additional upstream genes and downstream targets of transcription factors may be isolated by coupled insertional mutagenesis/enhancer trap screens and the numerous strategies available to identify and analyze specific protein–DNA complexes and their role(s) in early embryo gene regulation. These and complementary genetic approaches will lead to the elucidation of critical genes required for early embryogenesis and the regulatory networks that control them.

ACKNOWLEDGMENTS

Thanks to my colleagues, including those in my lab, who shared their most recently published and unpublished results. Work from the author's lab was supported by grants from Rhône-Poulenc Agrochimie, the Texas Advanced Research Program (No. 010366-038) and the U.S. Department of Agriculture Competitive Grants Program (No. 91-37304-6703).

REFERENCES

Allen, R.D., Trelease, R.N., and Thomas, T.L. (1988). Regulation of isocitrate lyase gene expression in sunflower. Plant Physiol. **86,** 527–532.

Allen, R.D., Bernier, F., Lessard, P.A., and Beachy, R.N. (1989). Nuclear factors interact with a soybean β-conglycinin enhancer. Plant Cell **1**, 623–631.

Amaya, E., Musci, T.J., and Kirschner, M.W. (1991). Expression of a dominant negative mutant of the FGF receptor disrupts mesoderm formation in Xenopus embryos. Cell **66**, 257–270.

Bäumlein, H., Wobus, U., Pustell, J., and Kafatos, F. (1986). The legumin gene family: Structure of a B-type gene of *Vicia faba* and a possible legumin gene specific regulatory element. Nucl. Acids Res. **14**, 2707–2720.

Bäumlein, H., Boerjan, W., Nagy, I., Panitz, R., and Wobus, U. (1991). Upstream sequences regulating legumin gene expression in heterologous transgenic plants. Mol. Gen. Genet. **225**, 121–128.

Bäumlein, H., Nagy, I., Villaroel, R., Inzé, D., and Wobus, U. (1992). *Cis*-analysis of a seed protein gene promoter: The conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene. Plant J. **2**, 233–239.

Bogue, M.A., Von der Haar, R.A., Nuccio, M.L., Griffing, L.R., and Thomas, T.L. (1990). Developmentally regulated expression of a sunflower 11S seed protein gene in transgenic tobacco. Mol. Gen. Genet. **221**, 49–57.

Bray, E.A., and Beachy, R.N. (1985). Regulation by ABA of β-conglycinin expression in cultured developing soybean cotyledons. Plant Physiol. **79**, 746–750.

Bustos, M.M., Guiltinan, M.J., Jordano, J., Begum, D., Kalkan, F.A., and Hall, T.C. (1989). Regulation of β-glucuronidase expression in transgenic tobacco plants by an A/T-rich, *cis*-acting sequence found upstream of a French bean β-phaseolin gene. Plant Cell **1**, 839–853.

Bustos, M.M., Begum, D., Kalkan, F.A., Battraw, M.J., and Hall, T.C. (1991). Positive and negative *cis*-acting DNA domains are required for spatial and temporal regulation of gene expression by a seed storage protein promoter. EMBO J. **10**, 1469–1479.

Chen, Z.-L., Pan, N.-S., and Beachy, R.N. (1988). A DNA sequence element that confers seed-specific enhancement to a constitutive promoter. EMBO J. **7**, 297–302.

Chuck, G., Robbins, T., Charanjit, N., Ralston, E., Courtney-Gutterson, N., and Dooner, H.K. (1993). Tagging and cloning of a petunia flower color gene with the maize transposable element *Activator*. Plant Cell **5**, 371–378.

Comai, L., Dietrich, R.A., Maslyar, D.J., Baden, C.S., and Harada, J.J. (1989). Coordinate expression of transcriptionally regulated isocitrate lyase and malate synthase genes in *Brassica napus* L. Plant Cell **1**, 293–300.

Cordewener, J., Booij, H., Van der Zandt, H., Van Engelen, F., Van Kammen, A., and De Vries, S. (1991). Tunicamycin-inhibited carrot somatic embryogenesis can be restored by secreted cationic peroxidase isoenzymes. Planta **184**, 478–486.

De Jong, A.J., Cordewener, J., Lo Schiavo, F., Terzi, M., Vandekerckhove, J., Van Kammen, A., and De Vries, S.C. (1992). A carrot somatic embryo mutant is rescued by chitinase. Plant Cell **4**, 425–433.

De Jong, A.J., Schmidt, E.D.L., and De Vries, S.C. (1993a). Early events in higher plant embryogenesis. Plant Mol. Biol. **22**, 367–377.

De Jong, A.J., Heidstra, R., Spaink, H.P., Hartog, M.V., Meijer, E.A., Hendriks, T., Lo Schiavo, F., Terzi, M., Bisseling, T., Van Kammen, A., and De Vries, S.C. (1993b). *Rhizobium* lipooligosaccharides rescue a carrot somatic embryo mutant. Plant Cell **5**, 615–620.

Dickinson, C.D., Evans, R.P., and Nielsen, N.C. (1988). *RY repeats* are conserved in the 5′-flanking region of legume seed protein genes. Nucl. Acids Res. **16**, 371.

Dietrich, R.A., Maslyar, D.J., Heupel, R.C., and Harada, J.J. (1989). Spatial patterns of gene expression in *Brassica napus* seedlings: Identification of a cortex-specific gene and localization of mRNAs encoding isocitrate lyase and a polypeptide homologous to proteinases. Plant Cell **1**, 73–80.

Dietrich, R.A., Radke, S.E., and Harada, J.J. (1992). Downstream DNA sequences are required to activate a gene expressed in the root cortex of embryos and seedlings. Plant Cell **4**, 1371–1382.

Dure, L., III, Crouch, M., Harada, J., Ho, T-H.D., Mundy, J., Quatrano, R., Thomas, T., and Sung, Z.R. (1989). Common amino acid sequence domains among the *lea* proteins of higher plants. Plant Mol. Biol. **12**, 475–486.

Fedoroff, N.V., and Smith, D.L. (1993). A versatile system for detecting transposition in *Arabidopsis*. Plant J. **3**, 273–289.

Finkelstein, R.R., Tenbarge, K.M., Shumway, J.E., and Crouch, M.L. (1985). Role of ABA in maturation of rapeseed embryos. Plant Physiol. **78**, 630–636.

Gatehouse, J.A., Evans, I.M., Croy, R.R.D., and Boulter, D. (1986). Differential expression of genes during legume seed development. Phil. Trans. Roy. Soc. Lond. Ser. B **314**, 367–384.

Gatz, C., Kaiser, A., and Wendenburg, R. (1991). Regulation of a modified CaMV 35S promoter by the Tn*10*-encoded Tet repressor. Mol. Gen. Genet. **227**, 229–237.

Goldberg, R.B. (1986). Regulation of plant gene expression. Phil. Trans. Roy. Soc. Lond. Ser. B **314**, 343–353.

Goldberg, R.B., Barker, S.J., and Perez-Grau, L. (1989). Regulation of gene expression during plant embryogenesis. Cell **56**, 149–160.

Goodrich, J., Carpenter, R., and Coen, E.S. (1992). A common gene regulates pigmentation pattern in diverse plant species. Cell **68**, 955–964.

Graham, I.A., Smith, L.M., Leaver, C.J., and Smith, S.M. (1990). Developmental regulation of expression of the malate synthase gene in transgenic plants. Plant Mol. Biol. **15**, 539–549.

Guerche, P., Tire, C., Grossi de Sa, F., De Clercq, A., Van Montagu, M., and Krebbers, E. (1990). Differential expression of the *Arabidopsis* 2S albumin genes and the effect of increasing gene family size. Plant Cell **2**, 469–478.

Guiltinan, M.J., Marcotte, W.R., and Quatrano, R.S. (1990). A leucine zipper protein that recognizes an abscisic acid response element. Science **250**, 267–270.

Harada, J.J., Baden, C.S., and Comai, L. (1988). Spatially regulated genes expressed during seed germination and postgerminative development are activated during embryogeny. Mol. Gen. Genet. **212**, 466–473.

Hattori, T., Vasil, V., Rosenkrans, L., Hannah, L.C., McCarty, D.R., and Vasil, I.K. (1992). The *Viviparous-1* gene and abscisic acid activate the *C1* regulatory gene for anthocyanin biosynthesis during seed maturation in maize. Genes Dev. **6**, 609–618.

Hughes, D.W., and Galau, G.A. (1988). Temporally modular gene expression during cotyledon development. Genes Dev. **3**, 358–369.

Jofuku, K.D., Okamuro, J.K., and Goldberg, R.B. (1987). Interaction of an embryo DNA binding protein with a soybean lectin gene upstream region. Nature **328**, 734–737.

Jordano, J., Almoguera, C., and Thomas, T.L. (1989). A sunflower helianthinin gene upstream sequence ensemble contains an enhancer and sites of nuclear protein interaction. Plant Cell **1,** 855–866.

Kamalay, J.C., and Goldberg, R.B. (1980). Regulation of structural gene expression in tobacco. Cell **19,** 935–946.

Kawagoe, Y., and Murai, N. (1992). Four distinct nuclear proteins recognize *in vitro* the proximal promoter of the bean seed storage protein β-phaseolin gene conferring spatial and temporal control. Plant J. **2,** 927–936.

Lam, E., and Chua, N.-H. (1991). Tetramer of a 21-base pair synthetic element confers seed expression and transcriptional enhancement in response to water stress and abscisic acid. J. Biol. Chem. **266,** 17131–17135.

Lelievre, J.-M., Oliveira, L.O., and Nielsen, N.C. (1992). 5′-CATGCAT-3′ elements modulate the expression of glycinin genes. Plant Physiol. **98,** 387–391.

Lessard, P.A., Allen, R.A., Bernier, F., Crispino, J.D., Fujiwara, T., and Beachy, R.N. (1991). Multiple nuclear factors interact with upstream sequences of the differentially regulated β-conglycinin genes. Plant Mol. Biol. **16,** 397–413.

Lessard, P.A., Allen, R.D., Fujiwara, T., and Beachy, R.N. (1993). Upstream regulatory sequences from two β-conglycinin genes. Plant Mol. Biol. **22,** 873–885.

Liang, P., and Pardee, A.B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science **257,** 967–971.

Liu, C.-m., Xu, Z.-h., and Chua, N.-H. (1993). Auxin polar transport is essential for the establishment of bilateral symmetry during early plant embryogenesis. Plant Cell **5,** 621–630.

Lopes, M.A., and Larkins, B.A. (1993). Endosperm origin, development, and function. Plant Cell **5,** 1383–1399.

Ludwig, S.R., Habera, L.F., Dellaporta, S.L., and Wessler, S.R. (1989). *Lc,* a member of the maize *R* gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the *myc*-homology region. Proc. Natl. Acad. Sci. USA **86,** 7092–7096.

Marcotte, W.R., Jr., Russell, S.H., and Quatrano, R.S. (1989). Abscisic acid response sequences from the Em gene of wheat. Plant Cell **1,** 969–976.

Mariani, C., De Beuckeleer, M., Truettner, J., Leemans, J., and Goldberg, R.B. (1990). Induction of male sterility in plants by a chimaeric ribonuclease gene. Nature **347,** 737–741.

McGurl, B., Pearce, G., Orozco-Cardenas, M., and Ryan, C.A. (1992). Structure, expression and antisense inhibition of the systemin precursor gene. Science **225,** 1570–1573.

Meinke, D.W. (1991). Perspectives on genetic analysis of plant embryogenesis. Plant Cell **3,** 857–866.

Mergny, J.L., Duval-Valentin, G., Nguyen, C.H., Perrouault, L., Faucon, B., Rougee, M., Montenay-Garestier, T., Bisagni, E., and Helene, C. (1992). Triple helix-specific ligands. Science **256,** 1681–1684.

Mundy, J., Yamaguchi-Shinozaki, K., and Chua, N.-H. (1990). Nuclear proteins bind conserved elements in the abscisic acid-responsive promoter of a rice *rab* gene. Proc. Natl. Acad. Sci. USA **87,** 1406–1410.

Nasrallah, J.B., and Nasrallah, M.E. (1993). Pollen–stigma signaling in the sporophytic self-incompatibility response. Plant Cell **5,** 1325–1335.

Oeda, K., Salinas, J., and Chua, N.-H. (1991). A tobacco bZip transcription activator (TAF-1) binds to a G-box-like motif conserved in plant genes. EMBO J. **10,** 1793–1802.

Pederson, T.J., Arwood, L.J., Spiker, S., Guiltinan, M.J., and Thompson, W.F. (1991). High mobility group chromosomal proteins bind to AT-rich tracts flanking plant genes. Plant Mol. Biol. **16,** 95–104.

Perez-Grau, L., and Goldberg, R.B. (1989). Soybean seed protein genes are regulated spatially during embryogenesis. Plant Cell **1,** 1095–1109.

Pysh, L.D., Aukerman, M.J., and Schmidt, R.J. (1993). OHP1: A maize basic domain/leucine zipper protein that interacts with Opaque2. Plant Cell **5,** 227–236.

Quatrano, R.S. (1987). The role of hormones during seed development. In Plant Hormones and Their Role in Plant Growth and Development, P.J. Davies, ed (Boston: Martinus Nijhoff Publishers), pp. 494–514.

Rivin, C.J., and Grudt, T. (1991). Abscisic acid and the developmental regulation of embryo storage proteins in maize. Plant Physiol. **95,** 358–365.

Salinas, J., Oeda, K., and Chua, N.-H. (1992). Two G-box–related sequences confer different expression patterns in transgenic tobacco. Plant Cell **4,** 1485–1493.

Schiavone, F.M., and Racusen, R.H. (1991). Regeneration of the root pole in surgically transected carrot embryos occurs by position-dependent, proximodistal replacement of missing tissues. Development **113,** 1305–1313.

Schindler, U., Beckmann, H., and Cashmore, A.R. (1992). TGA1 and G-box binding factors: Two distinct classes of Arabidopsis leucine zipper proteins compete for the G-box–like element TGACGTGG. Plant Cell **4,** 1309–1319.

Schmidt, R.J., Ketudat, M., Aukerman, M.J., and Hoschek, G. (1992). Opaque-2 is a transcriptional activator that recognizes a specific target site in 22-kD zein genes. Plant Cell **4,** 689–700.

Shotwell, M.A., and Larkins, B.A. (1989). The biology of seed storage proteins. In The Biochemistry of Plants, Vol. 15, A. Marcus, ed (New York: Academic Press), pp. 297–345.

Skriver, K., and Mundy, J. (1990). Gene expression in response to abscisic acid and osmotic stress. Plant Cell **2,** 503–512.

Sterk, P., Booij, H., Schellekens, G.A., Van Kammen, A., and De Vries, S.C. (1991). Cell-specific expression of the carrot EP2 lipid transfer protein gene. Plant Cell **3,** 907–921.

Thomas, T.L., Vivekananda, J., and Bogue, M.A. (1991). ABA regulation of gene expression in embryos and mature plants. In Abscisic Acid: Physiology and Biochemistry, W.J. Davies and H.G. Jones, eds (Oxford: Bios Scientific Publishers), pp. 125–135.

Van Engelen, F.A., and De Vries, S.C. (1992). Extracellular proteins in plant embryogenesis. Trends Genet. **8,** 66–70.

Van Engelen, F.A., Hartog, M.V., Thomas, T.L., Taylor, A., Sturm, A., Van Kammen, A., and De Vries, S.C. (1993). The carrot secreted glycoprotein gene EP1 is expressed in the epidermis and has sequence homology to *Brassica* S-locus glycoproteins. Plant J., in press.

Van Gelder, R.N., von Zastrow, M.E., Yool, A., Dement, W.C., Barehas, J.D., and Eberwine, J.H. (1990). Amplified RNA synthesized from limited quantities of heterogenous cDNA. Proc. Natl. Acad. Sci. USA **87,** 1663–1667.

Vellanoweth, R.L., and Okita, T.W. (1993). Regulation of expression of wheat and rice seed storage protein genes. In Control of Plant

Gene Expression, D.P.S. Verma, ed (Boca Raton, FL: CRC Press, Inc.), pp. 377–392.

Vilardell, J., Goday, A., Freire, M.A., Torrent, M., Martinez, M.C., Torne, J.M., and Pagès, M. (1990). Gene sequence, developmental expression and protein phosphorylation of RAB17 in maize. Plant Mol. Biol. **14,** 423–432.

Vivekananda, J., Drew, M.C., and Thomas, T.L. (1992). Hormonal and environmental regulation of the carrot *lea*-class gene *Dc3*. Plant Physiol. **100,** 576–581.

Voelker, T., Sturm, A., and Chrispeels, M.J. (1987). Differences in expression between two seed lectin alleles obtained from normal and lectin-deficient beans are maintained in transgenic tobacco. EMBO J. **6,** 3571–3577.

Wang, Z., and Brown, D.D. (1991). A gene expression screen. Proc. Natl. Acad. Sci. USA **88,** 11505–11509.

Weintraub, H., Davis, R., Tapscott, S., Thayer, M., Krause, M., Benezra, R., Blackwell, T.K., Turner, D., Rupp, R., Hollenberg, S., Zhuang, Y., and Lassar, A. (1991). The *myo*D gene family: Nodal point during specification of the muscle cell lineage. Science **251,** 761–766.

West, M.A.L., and Harada, J.J. (1993). Embryogenesis in higher plants: An overview. Plant Cell **5,** 1361–1369.

Wilde, R.J., Shufflebottom, D., Cooke, S., Jasinska, I., Merryweather, A., Berl, R., Brammar, W.J., Bevan, M.J., and Schuch, W. (1992). Control of gene expression in tobacco cells using a bacterial operator-repressor system. EMBO J. **11,** 1251–1259.

Zimmerman, J.L. (1993). Somatic embryogenesis: A model for early development in higher plants. Plant Cell **5,** 1411–1423.